

Метод построения общего домена радарных и оптических изображений дистанционного зондирования Земли для их совмещения

А.Н. Борисов¹, В.В. Мясников¹, В.В. Сергеев¹

¹ Самарский национальный исследовательский университет имени академика С.П. Королёва;
443086, Россия, г. Самара, Московское шоссе, д. 34

Аннотация

В работе предлагается метод построения общего домена для цифровых изображений дистанционного зондирования Земли, полученных двумя типами сенсоров: мультиспектральным оптическим сенсором и радаром с синтезированной апертурой. В общем домене изначально разнородные изображения должны обладать достаточным сходством, чтобы их совмещение могло выполняться методами совмещения обычных цифровых изображений. Задача построения общего домена рассмотрена в виде задачи одновременного обучения двух нейронных сетей U-Net. Предложенный метод сравнивается с существующим методом на основе генеративной нейронной сети pix2pix, выполняющей преобразование из радарного домена в оптический. Приводятся результаты совмещения изображений широко известным алгоритмом SIFT, современными алгоритмами RoMa и DeDoDe, и алгоритмом RIFT2, предназначенным для сопоставления разнородных изображений. Экспериментальные исследования демонстрируют однозначное преимущество предложенного метода: при использовании предложенного метода в комбинации с обученной сетью RoMa успешно совмещено 99,9% изображений со средней ошибкой отображения углов 0,84 пикселя; при использовании метода pix2pix в комбинации с DeDoDe успешно совмещено 98% изображений со средней ошибкой отображения углов 4,3 пикселя.

Ключевые слова: цифровые изображения дистанционного зондирования Земли, мультиспектральные изображения, радарные изображения, общий домен, преобразование домена, совмещение изображений.

Цитирование: Борисов, А.Н. Метод построения общего домена радарных и оптических изображений дистанционного зондирования Земли для их совмещения / А.Н. Борисов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. – 2026. – Т. 50, № 3. – 1980. – DOI: 10.18287/COJ1980.

Citation: Borisov AN, Myasnikov VV, Sergeev VV. Common domain construction method for radar and optical Earth remote sensing images matching. Computer Optics 2026; 50(3): 1980. DOI: 10.18287/COJ1980.

Введение

В современной мировой экономике одним из ключевых факторов для ее развития является информация, ее актуальность и полнота. И для многих отраслей источником такой информации являются средства космического дистанционного зондирования Земли (ДЗЗ). Центральное место здесь занимают спутники с оптическими (полноцветные, мультиспектральные, гиперспектральные) и радарными сенсорами (РСА – радар с синтезированной апертурой), количество которых постоянно растет по мере вовлечения в процесс ДЗЗ различных стран и крупных коммерческих компаний. Естественным процессом при получении, интерпретации и понимании информации является объединение данных с различных спутников, что невозможно без качественного совмещения (и координатной привязки) цифровых снимков. А по мере роста их числа и увеличения объемов получаемых данных задача совмещения цифровых изображений, полученных от сенсоров, использующих различные физические принципы регистрации информации, снимков разного пространственного и/или спектрального разрешения, снятых в разные моменты времени, становится все более актуальной.

Указанная задача совмещения цифровых изображений, называемая также задачей регистрации, часто возникала в приложениях фотограмметрии и смежных направлениях технического зрения [1, 2]: нахождение соответствующих точек на различных изображениях одного и того же объекта является ключевым шагом для построения/реконструкции 3D-модели регистрируемого объекта [3], решения задач локализации и картографирования [4], обнаружения изменений [5], слияния изображений [6] и др.

Развитие цифровых средств регистрации изображений различной физической природы (радарные, мульти- и гиперспектральные, видео), развитие персональных мобильных устройств и платформ размещения средств регистрации, высокопроизводительных средств обработки данных привело к взрывному росту решений в области компьютерного зрения в целом и методов корегистрации цифровых изображений ДЗЗ в частности. В результате методы совмещения прошли значительный путь от простых корреляционных методов и методов отождествления краев [7] до методов попиксельного совмещения (dense matching) с использованием глубоких нейронных сетей [1, 2].

При этом, несмотря на значительное разнообразие методов совмещения обычных (RGB) изображений, развитие методов совмещения для изображений, получаемых различными сенсорами (мульти- и гиперспектральными, PCA), остается сложной задачей в силу существенных различий между подобными изображениями.

Проблема точной корегистрации наиболее актуальна для оптических и радарных изображений. Оптические изображения отличаются удобством человеческой интерпретации и содержат несколько спектральных каналов (за исключением снимков одноканальных панхроматических сенсоров некоторых спутников), однако качество получаемых изображений зависит от погодных условий. Радарные снимки не зависят ни от погоды, ни от времени суток на месте съемки, однако страдают от спекл-шума.

В последнее время появился подход на основе использования мультимодальных и/или фундаментальных моделей ИИ [8, 9], использующих совместно радарные и оптические изображения для улучшения качества анализа. Очевидно, подобный анализ требует точной корегистрации изображений.

Современные методы совмещения оптических и PCA-изображений дистанционного зондирования, основанные на глубоком обучении, можно условно поделить на две категории [10].

Первый класс методов использует глубокое обучение для построения дескрипторов, применимых для сопоставления оптических и PCA-изображений: MatchosNet [11], PLISA [12] и др. *Методы второго класса* приводят изображения одного сенсора (исходного домена) к виду данных второго сенсора (целевого домена). Подобное преобразование между доменами, напоминающее конвертацию стилей изображений [13], типично для средств генеративного искусственного интеллекта [14, 15]. Для совмещения оптических и радарных изображений подобные решения на основе генеративно-состязательных сетей (generative adversarial networks GANs) применялись в работах [16, 17], похожие подходы к совмещению применялись и для снимков иных доменов [18].

Оба подхода демонстрируют достаточно высокие показатели качества, но имеют и недостатки. Общим является недостаточная универсальность подходов: при изменении свойств одного из изображений (например, числа каналов и/или их спектрального расположения) задачи создания дескрипторов (детекторов и описателей) или построения междоменного преобразования приходится решать полностью заново.

Дополнительным недостатком стандартного подхода со сменой домена является необходимость реконструировать специфические особенности изображений целевого домена из изображений исходного домена. Например, реконструировать спекл-шум, присущий PCA-изображениям, из изображений оптического сенсора или реконструировать цвет по данным PCA-изображения.

В данной работе мы предлагаем иной подход, который также использует преобразования входных оптических и PCA-изображений, но не требует полной смены их домена на альтернативный (это устраняет второй из указанных выше недостатков). Идея предлагаемого решения заключается в нахождении для оптических и PCA-изображений некоторого «общего» содержимого. То есть такого изображения, которое максимальным образом сохраняло бы общую информацию, присущую обоим изображениям, устраняя информацию о различиях. В некотором смысле находило «пересечение» между изображениями оптического и PCA-сенсоров.

На получаемые таким образом из оптических и/или радарных данных изображения мы далее ссылаемся как на *изображения нового – общего – домена*. Поскольку итоговый формат (число каналов) изображений общего домена может быть задан априори, их использование для решения задачи совмещения снимков оказывается чрезвычайно удобным и универсальным. А именно, мы можем либо использовать готовый метод совмещения (используемый для изображений с тем же числом каналов), либо обучить такой метод на новых данных – множестве изображений общего домена, получаемых для снимков исходных доменов (оптических и PCA-снимков).

В данной работе для совмещения мы использовали метод RoMa [19]. Как показали эксперименты, он без существенного изменения качественных показателей может применяться как в предобученном состоянии, так и с обучением на изображениях общего домена. В результате, при появлении снимков новых доменов можно не решать заново задачу обучения модели, выполняющей совмещение изображений, ограничиваясь задачей построения преобразования в общий домен.

Основной вклад работы следующий.

1. Предложен метод построения общего домена по обучающим данным: парам геометрически совмещенных снимков исходных доменов (в рамках данной работы – оптического и радарного сенсоров). Метод основан на совместном обучении нескольких нейронных сетей в условиях отсутствия требуемых выходов – искомым изображениям общего домена.
2. Предложено проводить совмещение пар цифровых снимков различных доменов (сенсоров) после преобразования в общий домен, т.к. изображения одной территории в построенном общем домене оказываются схожими, при этом «разнообразие» исходных данных в новом домене в основном сохраняется.
3. Представлены результаты экспериментов, демонстрирующие преимущество предлагаемого подхода совмещения оптических и радарного изображения с использованием общего домена. Показано, что совмещение изображений общего домена может проводиться существующими state-of-the-art нейросетевыми моделями DeDoDe и RoMa, обученными для совмещения обычных изображений, с достаточно высокой точностью. Наилучшее качество совмещения достигается при использовании модели RoMa, обученной на изображениях общего домена.

Работа построена следующим образом. В первом параграфе представлена краткая классификация существующих методов совмещения цифровых изображений в целом и совмещения оптических и РСА-изображений в частности. В этом же параграфе дана информация о доступных наборах данных для рассматриваемой задачи. Во втором параграфе представлено детальное описание предлагаемого метода, включая общую структуру решения (pipeline), конкретизацию используемых нейросетевых архитектур, а также описание используемых критериев обучения. Третий параграф содержит описание и результаты проведенных экспериментов по оценке эффективности предложенного метода совмещения оптических и РСА-изображений с использованием общего домена, дано сравнение решения с существующими state-of-the-art прототипами. В завершение работы представлено заключение и список использованных источников.

1. Современное состояние исследований

Несмотря на то, что задача совмещения снимков ДЗЗ может рассматриваться как частный случай задачи корегистрации изображений [1, 2], для совмещения снимков с различными физическими принципами их формирования, таких как оптические и РСА, потребовалась разработка и модифицированных, и совершенно новых решений. Ниже мы даем краткий обзор основных подходов к совмещению оптических и РСА-снимков с учетом принятой их классификации в последних релевантных работах [2, 10, 20, 21].

1.1. Методы и подходы для совмещения оптических и РСА-снимков

Методы на основе сопоставления с шаблоном (template-based methods, area-based methods) выполняют совмещение изображений в процессе решения оптимизационной задачи. Критерием задачи выступает агрегированная величина, составляющие которой – значения (рас)согласования пар соответствующих фрагментов изображений. Прямыми примерами подобного типа совмещения являются методы совмещения на основе максимизации взаимной информации [22] или нормализованной корреляции [23].

Современные методы этого класса дополнительно преобразуют фрагменты изображений в некоторое описание, формирующее вектор признаков фрагмента, и оценивают величину (рас)согласования фрагментов в пространстве этого описания. Примерами подобных описаний являются описатели HOG [24] и SFOG [25]. Очевидным недостатком методов этого класса (при их прямой реализации) является высокая вычислительная сложность, вызванная необходимостью перебора большого числа фрагментов на изображениях для расчета величин (рас)согласованности.

Методы совмещения изображений на основе сопоставления особых точек появились в результате развития методов сопоставления с шаблоном с целью устранения их основного недостатка. Радикального сокращения перебора удалось достичь за счет уменьшения числа потенциальных фрагментов-претендентов. В качестве них стали использовать только те, что содержали особые точки – это точки, которые устойчиво детектируются даже при наличии определенных преобразований изображений (как геометрических, так и яркостных). Примерами особых точек являются углы, специфические границы, центры замкнутых контуров и т.п.

Подход совмещения изображений с использованием особых точек обычно состоит из трех этапов. На первом этапе алгоритм обнаружения особых точек – детектор – на каждом из сопоставляемых изображений обнаруживает множества особых точек. На втором алгоритм описания особых точек каждой точки (а точнее, ее локальной окрестности/фрагменту) сопоставляет числовой вектор – дескриптор, описывающий ее свойства (свойства этой окрестности). Наконец, на третьем этапе алгоритм совмещения особых точек выполняет сравнение дескрипторов и формирует пары соответствующих друг другу точек на двух изображениях. В самом простом случае точки считаются соответствующими, если расстояние (евклидово, косинусное или др.) между их дескрипторами не превышает порога. Итоговое геометрическое преобразование между сопоставляемыми изображениями (в наиболее типичном случае описываемое матрицей проективного преобразования) оценивается по множеству пар соответствующих точек с алгоритмом RANSAC [26] или подобным ему. Наиболее известным методом совмещения обычных изображений по особым точкам является SIFT [27].

Для совмещения оптических и РСА-снимков изложенный подход на основе особых точек также эффективно применяется. Однако в силу различающихся принципов формирования изображений он потребовал изменения алгоритмов обнаружения и описания особых точек с учетом различных свойств оптических и РСА-изображений. Примерами методов совмещения на основе особых точек являются OS-SIFT [28] и PSO-SIFT [29]. Методы RIFT [30] and RIFT2 [31] также относятся к этой категории, хотя и не предназначались для работы с РСА-изображениями. Дополнительно следует отметить, что все перечисленные методы имеют ограничения на максимальное число каналов снимков: оно не превышает трех. И это делает невозможным использование мультиспектральных и/или гиперспектральных снимков ДЗЗ.

Интенсивное развитие в последнее десятилетие *методов глубокого обучения* существенным образом отразилось и на методах сопоставления снимков. Вместо детерминированных алгоритмов обнаружения и описания особых точек и фрагментов изображений, разрабатываемых с учетом опыта людей-экспертов, существенно эффективнее оказались обучаемые алгоритмы (глубоких нейронных сетей), настраиваемые по данным конкретных доменов. Использование глубокого обучения позволило не только адаптировать детекторы

и описатели к конкретным условиям наблюдения или доменам, но и позволило обеспечить сближение результатов конструируемых детекторов и описателей при совмещении разнородных изображений ДЗЗ. Примеры использования методов совмещения однородных изображений для совмещения изображений ДЗЗ можно найти в [32, 33]. Примерами методов, разработанных специально для решения задачи совмещения разнородных изображений ДЗЗ, являются ACAMatch [10], MatchosNet [11] и CAMM-net [33].

В работе [34], опубликованной авторами, сопоставление разнородных изображений выполняется с помощью комбинации метода главных компонент [35], выполняющего приведение многоканальных изображений ДЗЗ к 3 каналам, и обученной на главных компонентах модели RoMa.

Другим нововведением в задаче совмещения изображений ДЗЗ, связанным с развитием методов глубокого обучения, оказалась возможность *перевода изображений из одного домена в другой* (англ.: domain translation). Например, при совмещении оптического и РСА-изображений первое из них преобразуется к виду радарного, что позволяет для обоих изображений использовать одинаковые детекторы и дескрипторы.

Источником подобных решений явились методы генеративного искусственного интеллекта, в частности, генеративно-состязательные сети (GAN) [14, 15]. Примером такого решения является работа [36], где оптическое изображение преобразуется к виду радарного снимка с последующим обучением сети MatchosNet совмещению реальных и сгенерированных радарных изображений. В работе [19] исследуется в том числе качество совмещения методами SIFT, SAR-SIFT [37] и PSO-SIFT реальных и сгенерированных оптических изображений (то есть использовалось преобразование PCA-изображения в оптическое). Интересный подход используется в [38], где модифицированный вариант CycleGAN [15] обучается таким образом, чтобы информация в последнем латентном слое обоих генераторов была сходна. Информация из латентных слоев затем используется в качестве дескрипторов особых точек.

В целом, подход с использованием преобразования одного из совмещаемых изображений ДЗЗ в альтернативный домен оказался достаточно эффективным и часто используется в последнее время в том числе в задачах обнаружения объектов [39] и выявления изменений [40, 41].

Данный подход имеет объективный недостаток, обусловленный генерацией смоделированного изображения вместо физически обоснованного изображения целевого домена. В результате преобразование мультиспектральных и гиперспектральных изображений в радарные изображения приводит к значительной потере данных, поскольку исходные изображения имеют значительно большее количество полос, чем изображения SAR. Обратное преобразование, напротив, приводит к значительной нехватке данных.

В отличие от подхода, основанного на преобразовании из одного исходного домена в другой, предложенный в данной статье подход основан на построении общего домена. Этот подход не требует от нейронной сети восстановления деталей, отсутствующих в исходном домене; то есть нейронной сети не нужно «изобретать» детали. В качестве альтернативы предложенному подходу мы рассматриваем хорошо известный метод *pix2pix* [14], выполняющий преобразование из домена радарных снимков в домен оптических снимков.

1.2. Наборы данных

Для проведения исследований по совмещению оптических и РСА-изображений могут использоваться следующие наборы данных.

- SEN1-2 [42] – часто используемый набор. Он содержит 282 384 пары совмещенных фрагментов изображений, полученных европейскими спутниками Sentinel-1 (PCA-платформа) и Sentinel-2 (мультиспектральный сенсор) по всему земному шару. В рамках данного набора используются только VV-поляризация снимков Sentinel-1 и только RGB-каналы Sentinel-2. Изображения хранятся в PNG-формате, размер изображений – 256×256 пикселей. Шкала квантования для каждого канала – [0, 255].
- SEN12-MS [43] – набор данных, отличающийся от набора SEN1-2 форматом хранения изображений. Для PCA-изображений Sentinel-1 представлены обе поляризации, для Sentinel-2 даны все 13 спектральных каналов. Кроме того, этот набор данных включает в себя маски классификации поверхности, полученные по данным MODIS. Итоговый размер набора данных – 180 662 триплета. Изображения хранятся в TIFF-формате, размер изображений – 256×256 пикселей. Шкала квантования для оптических изображений – [0, 65 535], значения отсчетов PCA-снимков хранятся в виде вещественных чисел.
- OsDataset [44] – набор данных от китайского PCA-спутника GaoFen-3, включающий в себя 2 673 изображения в паре с оптическими RGB-изображениями из Google Earth. PCA-изображения представлены в оттенках серого. Изображения хранятся в PNG-формате, размер изображений – 512×512 пикселей. Шкала квантования – [0, 255].
- MultiSenGE [45] – мультимедийный набор данных, включающий в себя совмещенные фрагменты снимков Sentinel-1 и Sentinel-2, полученных с июля по ноябрь 2020 г. над востоком Франции. Набор также включает маски сегментации земной поверхности высокого разрешения. В отличие от набора SEN12-MS цифровые изображения подвергались предварительной обработке: для PCA-снимков выполнено подавление импульсного шума и перевод значений в линейную шкалу; для оптических снимков исключены каналы 1, 9 и 10, имеющие низкое разрешение (итоговое количество каналов – десять). Набор данных содержит 72 033 фрагмента снимков Sentinel-2 и 1 012 227

фрагментов Sentinel-1, соответствующих 8 157 различным локациям. Изображения хранятся в TIFF-формате, размер изображений – 256×256 пикселей. Шкала квантования для оптических изображений – $[0, 65\ 535]$, значения отсчетов РСА-снимков хранятся в виде вещественных чисел.

Для экспериментов был выбран набор MultiSenGe по причине наличия мультивременных серий снимков и удобного представления радарных снимков, не требующего дополнительной предобработки.

2. Метод совмещения разнородных снимков ДЗЗ

2.1. Описание метода

Общая структура предлагаемого метода представлена на рис. 1. На вход метода поступает два изображения размерами $W_1 \times H_1$ и $W_2 \times H_2$ с K_1 и K_2 каналами соответственно. Каждое из этих изображений преобразуется в новое изображение с предопределенным числом каналов K . В отличие от существующих методов на основе смены домена мы не приводим изображение одного исходного домена к другому исходному домену. Вместо этого мы переводим оба изображения в некоторый новый домен с предопределенным числом каналов K , но с сохранением их линейных размеров. В этом новом домене – *общем домене* – «общее» содержимое исходных изображений сохраняется, а различия между ними нивелируются.

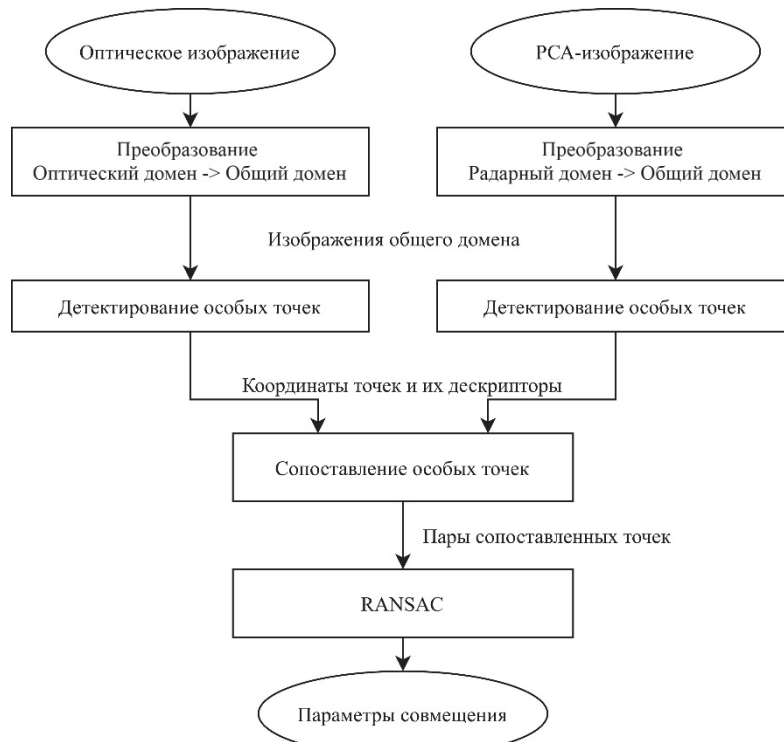


Рис. 1. Общая схема предлагаемого метода

Ниже в подпараграфе 2.2 приведено детальное описание предложенного метода: вводится понятие общего домена, формализуется задача построения преобразования к общему домену и приводятся варианты ее решения. В подпараграфе 2.3 описываются способы совмещения изображения.

2.2. Определение общего домена и его построение

Под *изображениями общего домена*, полученными из пары изображений разных доменов (в частности, оптических и радарных), мы понимаем пару цифровых изображений, которые удовлетворяют следующим требованиям: линейные размеры H, W изображений общего домена совпадают с линейными размерами изображений исходных доменов; изображения общего домена имеют наперед заданное и одинаковое число каналов (ниже мы обозначает его как K); изображения общего домена для геометрически совмещенных снимков одинакового линейного размера различных доменов должны быть схожи (близки по некоторой мере); изображения общего домена (и пара соответствующих преобразований) не являются вырожденными, то есть в них «сохраняется» значительная доля «информации» снимков исходных доменов.

Последние два требования не являются до конца определенными, также они противоречивы. В частности, требованию (третьему) схожести изображений общего домена легко удовлетворить, построив вырожденные преобразования, которые формирует изображения в виде константного поля. Четвертое требование как раз и предназначено для исключения подобных ситуаций.

Одной из возможных, но не единственной конкретизацией последнего требования является возможность достаточно точной реконструкции снимков исходного домена по изображениям из общего домена. Ниже мы рассмотрим такую конкретизацию последнего требования, как и более простую для задачи построения соответствующего преобразования. Аналогично, конкретизация третьего требования, как и архитектуры конструируемого преобразования, также представлены ниже.

Для последующего изложения нам потребуется некоторая формализация. Обозначим как

$$S = \{(x_i, y_i) | x_i \in \mathbb{R}^{W \times H \times K_X}, y_i \in \mathbb{R}^{W \times H \times K_Y}\}_{i=1}^N$$

набор геометрически совмещенных пар цифровых снимков $W \times H$, полученных от систем ДЗЗ с K_X и K_Y каналами ($K_X, K_Y \in \mathbb{N}$) соответственно. Для определенности будем считать, что изображения x_i – это РСА-изображения домена X , а y_i – оптические изображения домена Y . Проблема построения преобразования снимков в общий домен может быть конкретизирована в виде оптимизационной задачи нахождения пары отображений $(\varphi_X^*, \varphi_Y^*)$ следующего вида:

$$(\varphi_X^*, \varphi_Y^*) = \underset{\substack{\varphi_X \in \mathbf{Y}_X, \\ \varphi_Y \in \mathbf{Y}_Y}}{\operatorname{argmin}} \mathbb{E}_{(x,y) \in S} [L(\varphi_X(x), \varphi_Y(y)) + \alpha F(x, \varphi_X(x), y, \varphi_Y(y))], \quad (1)$$

где $L: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}$ – функция, характеризующая меру схожести двух изображений в общем домене; $F(\cdot)$ – вещественнозначная функция четырех аргументов $\mathbb{R}^{W \times H \times K}$, характеризующая вырожденность изображений $\varphi_X(x), \varphi_Y(y)$ общего домена относительно изображений x, y исходных доменов; $\mathbf{Y}_*$ – подмножества/классы отображений вида $\varphi_*: \mathbb{R}^{W \times H \times K_*} \rightarrow \mathbb{R}^{W \times H \times K}$, по которым производится поиск решения, $\alpha \in \mathbb{R}_+$ – весовой коэффициент.

Заметим, что есть два существенных отличия предлагаемого подхода с общим доменом от существующих решений с преобразования одного из сопоставляемых изображений к альтернативному исходному домену.

Первое: при преобразовании к существующему домену требуемый результат преобразования (изображение альтернативного домена) уже существует и может использоваться для обучения преобразования. При использовании общего домена требуемый результат оказывается неизвестен. В результате и «правильный» выход ГНС, и итоговое преобразование снимков в общий домен определяются в результате обучения. И результат в значительной степени определяется видом критериев L и F , наряду с выбранными классами преобразований $\mathbf{Y}_*$ (то есть видом используемых ГНС) и методами их обучения.

Вторым принципиальным отличием при использовании общего домена является отсутствие необходимости реконструировать специфические особенности изображений одного домена из изображений другого домена. Например, реконструировать спекл-шум, присущий радарным изображениям, из изображений оптического сенсора, или реконструировать цвет по данным радарного изображения. Очевидно, что исключение подобных неестественных свойств конструируемых преобразований является позитивным шагом, позволяющим (как покажут результаты экспериментов в третьем параграфе) улучшить качественные результаты решения задачи совмещения снимков.

Сходная оптимизационная задача для спектрального согласования оптических изображений была рассмотрена в [46, 47]. Было показано, что достаточно хорошее (линейное) решение можно найти аналитически, наилучшие результаты достигаются реализацией преобразований из $\mathbf{Y}_*$ в виде нейронных сетей.

Требование «невыврожденности» изображений общего домена означает, что изображения не вырождаются в (почти) постоянные области (при условии, что исходные изображения не являются таковыми). С другой стороны, требование сходства изображений в общем домене означает, что такие изображения должны иметь общий набор пространственных особенностей, а средняя яркость должна быть одинаковой. Следовательно, при построении функции L необходимо учитывать как различия в яркости изображений, так и сохранение пространственных особенностей.

В работе [48] рассматривается применение различных функций потерь в задаче реконструкции изображений. Было показано, что включение индекса структурного сходства ($SSIM$) [49] в функцию потерь позволяет получать лучшие результаты по сравнению с использованием только MAE/MSE как функции потерь.

Как следствие, в качестве функции L , характеризующей меру схожести двух изображений $\varphi_X(x), \varphi_Y(y)$ общего домена, мы используем следующую (β, γ – неотрицательные вещественные коэффициенты):

$$L(f_X(x), f_Y(y)) = \beta MSE(\varphi_X(x), \varphi_Y(y)) + \gamma DiSSIM(\varphi_X(x), \varphi_Y(y)). \quad (2)$$

Второе слагаемое определяется следующим образом:

$$DiSSIM(\varphi_X(x), \varphi_Y(y)) = 1 - SSIM(\varphi_X(x), \varphi_Y(y)),$$

где $SSIM$ – индекс структурного сходства изображений. Для его вычисления мы использовали окна $w \times w$ размером $w = 5$, усредняя на большом многоканальном изображении результаты по всем возможным окнам и каналам. Стоит отметить, что в работе [48] рассматривалось также использование $MS-SSIM$ и MAE , но в настоящей работе в ходе экспериментов лучшие результаты были получены при использовании MSE и $SSIM$.

При определении составляющей критерия $F(\cdot)$, который характеризует вырожденность изображений $\varphi_X(x), \varphi_Y(y)$ общего домена относительно исходных изображений x, y , мы рассматриваем два варианта, существенно различающихся и архитектурой, и вычислительными затратами.

Первый вариант соответствует уже упомянутой выше конкретизации последнего требования к общему домену в виде возможности достаточно качественной реконструкции снимков исходных доменов по изображениям общего домена. В этом случае задача (1) может рассматриваться как обучение пары автоэнкодеров [50]. Обозначим энкодеры как $\varphi_X(x)$ и $\varphi_Y(y)$ и декодеры, реализующие обратное преобразование, как $\varphi_X^{-1}(\cdot)$ и $\varphi_Y^{-1}(\cdot)$. В этом случае F можно определить как сумму двух функций потерь:

$$L_F(a, b) = \text{MSE}(a, b) + \text{DiSSIM}(a, b), \\ F(x, \varphi_X(x), y, \varphi_Y(y)) = L_F(x, \varphi_X^{-1}(\varphi_X(x))) + \eta L_F(y, \varphi_Y^{-1}(\varphi_Y(y))). \quad (3)$$

Здесь $L_F: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}$ – функции, характеризующие меру схожести двух изображений заданного домена, и $\eta \in \mathbb{R}_+$ – весовой коэффициент, определяющий вклад значений от каждого из доменов.

Учитывая, что процесс реконструкции изображений исходных доменов по изображениям общего домена архитектурно и вычислительно сложен, мы также рассматриваем и более простой (второй) вариант.

Второй вариант показателя F основан на измерении сходства полутоновых изображений для обеспечения невырожденности преобразований. Для радарного изображения соответствующее ему полутоновое изображение извлекается из его образа в общем домене. Для оптического изображения соответствующее ему полутоновое изображение получается из RGB-каналов (или других каналов с наивысшим разрешением). В итоге, вторая версия показателя F выглядит следующим образом.

$$F(x, f_X(x), y, f_Y(y)) = \text{DiSSIM}(\mu(\varphi_X(x)), \mu(y|_{HR})), \quad (4)$$

где $\mu: \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}^{W \times H \times 1}$ – функция, усредняющая все каналы изображения, $y|_{HR}$ – результат выделения каналов высокого разрешения из оптического изображения.

Вторая версия показателя F устраняет необходимость построения обратных преобразований для обоих исходных доменов и, как правило, требует меньшего количества обученных нейронных сетей.

Стоит отметить, что изначально F рассматривалась в следующем варианте:

$$F(x, \varphi_X(x), y, \varphi_Y(y)) = \lambda_1 \text{DiSSIM}(\mu(\varphi_X(x)), \mu(y|_{HR})) + \lambda_2 \text{DiSSIM}(\mu(\varphi_Y(y)), \mu(y|_{HR})),$$

однако при значениях $\lambda_2 > 0$ показатели качества ухудшались. Данный факт можно объяснить исходя из следующих соображений:

- если $\lambda_2 = 0$, значение (1) зависит только от сходства изображений в общем домене и от структурного сходства радарного изображения в общей домена и оптического изображения. Можно сказать, что $\varphi_X(x)$ обучается восстановлению особенностей оптического изображения по радарному изображению. Поскольку полное восстановление невозможно и часть деталей все же теряется, $\varphi_Y(y)$ обучается упрощению оптического изображения с потерей тех деталей, которые не могут быть реконструированы $\varphi_X(x)$;
- если $\lambda_2 > 0$, потеря любых деталей будет увеличивать значение (1), как следствие, $\varphi_Y(y)$ будет отбрасывать меньше деталей, что увеличит расхождение между $\varphi_X(x)$ и $\varphi_Y(y)$.
- В результате мы имеем следующее формальное описание преобразований в предлагаемом методе:
- x и y – совмещенные оптическое и радарное изображение (снимки/изображения исходных доменов);
- $\tilde{x} = \varphi_X(x)$ и $\tilde{y} = \varphi_Y(y)$ – изображения общего домена;
- $\hat{x} = \varphi_X^{-1}(\tilde{x})$ и $\hat{y} = \varphi_Y^{-1}(\tilde{y})$ – реконструированные изображения в исходных доменах.

Обучаемыми нейросетями в общем случае являются: $\varphi_X(\dots)$, $\varphi_Y(\dots)$, $\varphi_X^{-1}(\dots)$ и $\varphi_Y^{-1}(\dots)$. Для построения преобразований $\tilde{x} = \varphi_X(x)$ и $\tilde{y} = \varphi_Y(y)$ изображений исходных доменов в общий домен использовалась сеть U-Net [51] с 5 уровнями. Сеть Efficient-Net-B0 [52] использовалась как кодирующая часть с целью улучшения формируемых описаний. В качестве активации последнего слоя сети используется сигмоида, приводящая выходные значения к интервалу (0, 1).

Для построения реконструирующих преобразований $\hat{x} = \varphi_X^{-1}(\tilde{x})$ и $\hat{y} = \varphi_Y^{-1}(\tilde{y})$ изображений из общего в исходные домены мы использовали аналогичную сеть с обратным числом входных и выходных каналов, но без использования сигмоиды на выходе сети.

2.3. Совмещение изображений общего домена

После преобразования изображений в общий домен их можно совместить с помощью методов сопоставления обычных изображений. Преимуществом этого подхода является наличие готовых реализаций таких методов в распространенных библиотеках компьютерного зрения. Наиболее известным методом сопоставления обычных изображений является SIFT [27].

Среди доступных современных (англ.: state-of-the-art) моделей нейронных сетей для сопоставления изображений выделяются модели DeDoDe и RoMa. Модель DeDoDe [53] – это метод на основе обнаружения особых точек. Она состоит из двух сетей: сети-детектора особых точек и сети-дескриптора, которая вычисляет описание признаков ключевых точек. Дескрипторы признаков сравниваются на основе их близости. Модель RoMa [19] – это метод попиксельного сопоставления (dense matching). Результатом метода является карта соответствий между всеми пикселями первого и второго изображений, а также оценка качества для каждой пары пикселей. Такой подход потенциально позволяет находить соответствия между изображениями в условиях сильных нелинейных искажений.

Стоит отметить, что координатные плоскости ортокорректированных изображений дистанционного зондирования связаны аффинным преобразованием. Использование попиксельного отображения для таких изображений излишне. Поэтому для оценки параметров преобразования мы выбираем несколько пар точек с наилучшим качеством.

3. Эксперименты

3.1. Используемые алгоритмы и наборы данных

В качестве альтернативного метода междоменного преобразования используется преобразование PCA-снимков в оптический домен нейросетью pix2pix [14].

В качестве методов, используемых непосредственно для совмещения изображений, используются методы SIFT [27], RIFT2 [31], DeDoDe [53] и RoMa [19]. Количество детектируемых на изображении особых точек ограничено 4096.

В качестве данных – оптических и PCA-изображений – для обучения и тестирования использовался набор данных MultiSenGE. Поскольку MultiSenGE для каждой области съемки содержит несколько радарных и оптических изображений, полученных в разное время, формирование обучающих пар выполнялось следующим образом: для каждого оптического изображения выбиралось радарное изображение для той же территории с ближайшей датой съемки. Полученный набор данных был разделен на обучающую, тестовую и валидационную выборки так, чтобы не допускать пересечений по территории (все пары снимков определенной территории находятся полностью в одном из наборов). Обучающая выборка состояла из 46 092 пар, валидационная – из 11 522 пар, тестовая – из 14 405 пар снимков. Значения отсчетов радарных изображений были переведены в логарифмическую шкалу.

Поскольку существующие программные решения для обучения генеративных сетей на момент проведения экспериментов поддерживали работу исключительно с RGB- или полутоновыми изображениями, был создан вспомогательный набор стандартных RGB-изображений, полученный следующим образом:

- радарные снимки были переведены в трехканальные изображения путем помещения VV-поляризации в красном канале и VH-поляризации в зеленом канале с последующим приведением к логарифмической шкале;
- оптические (мультиспектральные) изображения были преобразованы в цветные изображения путем отбора RGB-каналов.

Для последующего приведения диапазона значений отсчетов снимков к единому интервалу $[0, 1]$ использовалась следующая процедура: для каждого типа изображений (оптических и радарных) вычислялись глобальные процентиля $p_{0.5}$ и $p_{99.5}$, после чего диапазон значений отсчетов изображений ограничивался интервалом $[p_{0.5}, p_{99.5}]$ с последующим линейным преобразованием в диапазон $[0, 1]$.

3.2. Обучение нейросетей

Эксперименты проводились на машине с AMD Ryzen 5600X, 32 GB ОЗУ и Nvidia GeForce RTX 3090.

В качестве фреймворка для машинного обучения использовался PyTorch 2.7. Реализация Unet взята из пакета segmentation-models-pytorch [66].

Для всех вариантов преобразования в общий домен $K_X = 2$, $K_Y = 10$, а итоговое число каналов $K = 3$. Это позволило использовать богатую номенклатуру методов совмещения изображений.

Использовались следующие общие параметры обучения: метод оптимизации – NAdam [55], базовый темп обучения – 2×10^{-4} , метод изменения темпа обучения – OneCycle [56], количество эпох – 512, размер батча – 32, для аугментации использовались случайные отражения и повороты на 90° . Для обучения с критерием (3) использовались следующие параметры критерия из формул (1)–(3): $\alpha = 1$, $\beta = 3$, $\gamma = 8$, $\eta = 1$. Для обучения с критерием (4) использовались следующие параметры критерия из формул (1)–(2): $\alpha = 16$, $\beta = 1$.

Для обучения pix2pix использовался код из репозитория [57]. Обучение pix2pix осуществлялось со стандартными параметрами без изменения кода. В качестве набора данных использовался набор цветных изображений, получение которого описано в подпараграфе 3.2.

Поскольку методы сопоставления изображений, основанные на нейронных сетях, могут быть адаптированы к конкретному домену, мы решили выполнить обучение для лучшей комбинации (метод, основанный на критерии (4) + RoMa). Обучение сети RoMa проводилось со стандартными параметрами, за исключением размера

изображений, установленного в 224×224 , с аугментациями в виде случайного сдвига до 64 пикселей, случайными поворотами на угол до 30° и масштабированием со случайным коэффициентом из интервала $[0,9; 1,1]$.

3.3. Показатели качества

В качестве мер, характеризующих сходство изображений, наряду с *SSIM*, используются следующие критерии:

$$RMSE(\tilde{x}, \tilde{y}) = \sqrt{MSE(\tilde{x}, \tilde{y})}, PSNR(\tilde{x}, \tilde{y}) = -10 \lg MSE(\tilde{x}, \tilde{y}).$$

Формула *PSNR* приведена с учетом того, что значения отсчетов сравниваемых изображений находятся в отрезке $[0; 1,0]$.

Оценка качества совмещения изображений ДЗЗ проводилась нами на примере решения задачи нахождения параметров/матрицы $T \in \mathbb{R}^{3 \times 3}$ проективного преобразования, связывающих координатные плоскости оптического и радарного изображений. Поскольку изображения в наборе данных уже сопоставлены, T является единичной матрицей.

Само проективное образование определяется следующим образом. Пусть $h(p)$ – оператор приведения пиксельных координат к гомогенным координатам, а $h^{-1}(p)$ – оператор приведения гомогенных координат к пиксельным:

$$h(p) = (x, y, 1 | p = (x, y))^T, h^{-1}(p_h) = (xw^{-1}, yw^{-1} | p_h = (x, y, w)^T).$$

Оператор проецирования $Pr(T, p)$ задает проективное преобразование:

$$Pr(T, p) = h^{-1}(T \cdot h(p)),$$

где $T \in \mathbb{R}^{3 \times 3}$ – матрица, задающая проективное преобразование, $p \in \mathbb{R}^2$ – координаты точки.

Матрица искомого проективного преобразования определялась по набору пар $P = \{(p_1, p_2)_k | p_1 \in P_X, p_2 \in P_Y\}$ соответствующих точек. Здесь $P_X = \{p^i \in \mathbb{R}^2\}$ – множество особых точек радарного изображения, P_Y – множество особых точек оптического изображения, найденных предлагаемым или альтернативными методами.

Для определения множества пар соответствующих точек мы использовали алгоритм RANSAC [26] с предельным числом итераций 2000. Заметим, что результат RANSAC – это не только матрица $\hat{T} \in \mathbb{R}^{3 \times 3}$, являющаяся оценкой истинной матрицы проективного преобразования, но также подмножество пар точек $P_+ = \{(p_1, p_2) | (p_1, p_2) \in P, \|Pr(\hat{T}, p_1) - p_2\| \leq 3\}$ (где $\|\cdot\|$ – евклидова норма), согласованных с матрицей $\hat{T}$ (т.е. пары точек, расстояние между которыми после проецирования не превышает 3 пикселей).

На основе данных обозначений определяются следующие показатели качества совмещения изображений.

Показатель MMA (mean matching accuracy) определяется как доля соответствующих точек, удовлетворяющих полученному проективному преобразованию:

$$MMA = |P_+|/|P|.$$

Показатель ACE отражает величину рассогласования угловых точек оптического и РСА-изображений после их геометрического согласования:

$$Corners = \{(0, 0), (0, 255), (255, 0), (255, 255)\}, ACE = \frac{1}{4} \sum_{c \in Corners} \|Pr(\hat{T}, c) - c\|,$$

где $\|\cdot\|$ – евклидова норма.

Показатель SR (success rate) определяется как доля изображений с $ACE < 40$ (в случае провала сопоставления ACE считается равным бесконечности).

Для оценки координатной точности сопоставления используется *показатель Δ_P* , определяемый как среднее расстояние между соответствующими точками:

$$\Delta_P = \frac{1}{|P|} \sum_{(p_1, p_2) \in P} \|Pr(T, p_1) - p_2\|.$$

Показатель CMR (correct matches rate) определяется как доля пар точек, для которых истинная координатная ошибка не превосходит 3 пикселей:

$$CMR = |P_C|/|P|, P_C \triangleq \{(p_1, p_2) | (p_1, p_2) \in P, \|Pr(T, p_1) - p_2\| \leq 3\}.$$

Важно отметить, что, в отличие от *MMA*, при расчете *CMR* используется истинная матрица проективного преобразования T .

3.4. Результаты экспериментов

На первом этапе наших экспериментов рассматривалось сходство изображений, полученных различными методами междоменного преобразования.

Примеры результатов различных междоменных преобразований приведены на рис. 2.

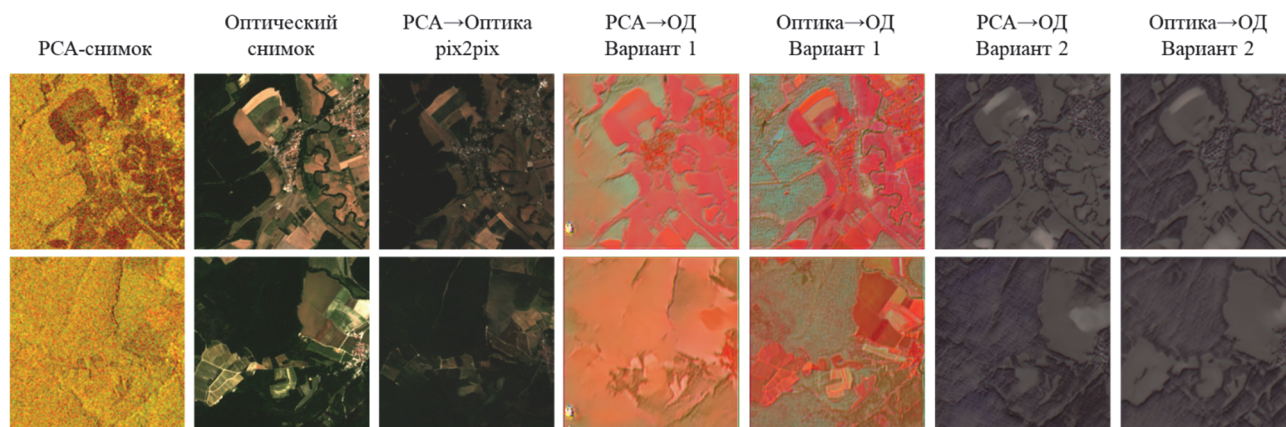


Рис. 2. Примеры преобразованных изображений

Столбцы 1-2 на этом рисунке содержат изображения без каких-либо доменных преобразований: цветное RGB-изображение, извлеченное из соответствующих каналов оптического изображения, и трехканальное изображение, полученное из радарного (красный канал – VV-поляризация, зеленый канал – VH-поляризация, синий канал занулен). Значения отсчетов полученных изображений приводились к отрезку $[0; 1]$ процедурой, описанной в подпараграфе 3.1 выше. Столбец 3 содержит результаты преобразования РСА-изображения в оптический домен обученной моделью pix2pix. Столбцы 4-5 и 6-7 содержат результаты преобразования оптических и РСА-изображений в общий домен, построенный по критериям (3) и (4) соответственно. Стоит отметить, что изображения из общего домена, построенного по критерию (3), имеют низкий контраст. Для улучшения контрастности применялась процедура нормализации, описанная в подпараграфе 3.2. Изображения из общего домена, построенного по критерию (4), обладают достаточным динамическим диапазоном и не требуют нормализации.

Средние значения показателей качества представлены в табл. 1.

Табл. 1. Значения показателей сходства изображений

Метод преобразования	RMSE	PSNR	SSIM
-	0,36	8,91	0,03
pix2pix	0,10	20,38	0,49
Вариант-1	0,13	18,66	0,53
Вариант-2	0,06	25,05	0,75

Из рассматриваемой таблицы видно, что сходство между оптическими и РСА-изображениями в общем домене значительно выше, чем между оптическими и псевдооптическими изображениями, сгенерированными из РСА-изображений моделью pix2pix.

На втором этапе экспериментов мы провели численное сравнение качества сопоставления изображений на основе критериев, изложенных в подпараграфе 3.4. Результаты экспериментов представлены в табл. 2-7. Для каждого рассмотренного метода доменного преобразования мы приводим результаты для всех четырех методов сопоставления признаков: SIFT, DeDoDe, RIFT2 и RoMa. Лучшее значение показателя качества выделено в таблице жирным цветом (в табл. 4 за лучшее значение взят результат DeDoDe, т.к. метод RoMa позволяет получить любое заданное количество точек).

Как видно, предложенный метод построения общего домена на основе критерия (4) обеспечивает значительно лучшие показатели качества совмещения изображений. Важно отметить, что результаты достигнуты при использовании методов сопоставления обычных изображений.

Табл. 2. Доля успешно совмещенных изображений (показатель SR)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	0	0,09	0,22	0,01
pix2pix	0,16	0,31	0,98	0,68
Вариант-1	0,03	0,29	0,98	0,61
Вариант-2	0,51	0,67	0,994	0,97

Также из полученных результатов видно, что модель RoMa обеспечивает наилучшую точность сопоставления изображений, хотя и уступает по показателю SR методу DeDoDe. С целью оценки возможного улучшения показателей модель RoMa была обучена на изображениях общего домена, построенного по критерию (4).

Результаты обученной модели RoMa (наряду с результатами других методов сопоставления изображений для метода на основе критерия (4)) приведены в табл. 8.

Табл. 3. Средняя ошибка совмещения изображений (показатель ACE)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	-	21,1	21,1	14,2
pix2pix	13,9	19,3	4,3	6,3
Вариант-1	12,8	19,0	4,1	4,7
Вариант-2	9,42	14,6	2,04	1,4

Табл. 4. Среднее количество пар точек (показатель $|P|$)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	2,7	38,8	168,1	4096
pix2pix	7,7	84,9	570,5	4096
Вариант-1	2,8	67,7	424,1	4096
Вариант-2	10,7	184,8	859,7	4096

Табл. 5. Средняя координатная ошибка сопоставленных точек (показатель Δ_p)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	125,3	49,2	89,7	144,1
pix2pix	77,7	31,1	14,8	56,5
Вариант-1	73,8	26,9	11,4	52,1
Вариант-2	29,9	16,3	4,4	7,7

Табл. 6. Среднее значение показателя MMA

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	0	0,31	0,13	0,33
pix2pix	0,18	0,30	0,37	0,46
Вариант-1	0,10	0,34	0,48	0,56
Вариант-2	0,66	0,32	0,64	0,85

Табл. 7. Средняя доля корректно сопоставленных точек (показатель CMR)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	0,005	0,10	0,04	0,004
pix2pix	0,25	0,15	0,42	0,35
Вариант-1	0,22	0,18	0,52	0,37
Вариант-2	0,57	0,26	0,72	0,87

Табл. 8. Показатели качества совмещения

Метод совмещения	Показатель качества					
	SR	ACE	$ P $	Δ_p	MMA	CMR
SIFT	0,51	9,42	10,7	29,9	0,66	0,57
RIFT2	0,67	14,6	184,8	16,3	0,32	0,26
DeDoDe	0,994	2,04	859,7	4,4	0,64	0,72
RoMa без обучения	0,97	1,4	4096	7,7	0,85	0,87
RoMa с обучением	0,999	0,84	4096	1,03	0,96	0,994

Полученные результаты позволяют заключить, что использование модели RoMa, обученной на изображениях общего домена, построенного с использованием критерия (4), обеспечивает наилучшее качество совмещения радарных и оптических снимков ДЗЗ с долей успешных сопоставлений 99,9 % и средней ошибкой совмещения изображений ACE менее 1 пикселя. При этом стоит отметить, что для других методов преобразования наилучшие показатели совмещения обеспечивает метод DeDoDe.

Заключение

В данной работе предложен метод построения общего домена для оптических и РСА-изображений дистанционного зондирования Земли. В отличие от существующих решений, основанных на преобразовании снимков

одного домена в другой, априорно заданный домен (например, PCA-изображения в псевдооптическое), общий домен не является доменом одного из источников данных, а конструируется на основании анализа изображений обоих исходных доменов. В общем домене удастся добиться как сходства изображений из разных исходных доменов, так и сохранения их основной информации, требуемой для успешного решения задачи совмещения снимков.

Предложена формализация проблемы построения общего домена в виде оптимизационной задачи, получено ее решения для двух частных критериев в виде соответствующих преобразований с использованием методов глубокого обучения и генеративных нейросетей.

В качестве прикладного применения преобразования к общему домену рассмотрена задача совмещения разнородных изображений ДЗЗ. Показано, что преобразование изображений к общему домену позволяет проводить совмещение изначально разнородных изображений с помощью методов, разработанных для совмещения обычных цифровых изображений.

Экспериментальные исследования проведены на наборе данных MultiSenGE, состоящем из пар мультиспектральных оптических и PCA-снимков. В качестве методов сопоставления изображений использовались методы SIFT, RIFT2, DeDoDe и RoMa. Результаты проведенных экспериментов продемонстрировали преимущество предлагаемого решения в качественном плане по сравнению решением, основанным на генерации псевдооптических снимков по PCA-данным моделью pix2pix, при этом нейросетевые методы DeDoDe и RoMa продемонстрировали высокую точность сопоставления изображений без переобучения на изображениях общего домена. Обучение модели RoMa на изображениях общего домена позволило достичь наилучших характеристик совмещения изображений (доля успешно совмещенных изображений – 0,999; средняя ошибка совмещения углов изображений – 0,84 пикселя).

Перспективным направлением развития предложенного метода является выбор как более эффективных составляющих оптимизационной задачи, определяющих общий домен, так и используемых генеративных сетей для построения соответствующих преобразований изображений между доменами.

Благодарности

Результаты исследования были получены при финансовой поддержке Российской Федерации в лице Минобрнауки России (Соглашение № 075-15-2024-558).

References

- [1] Verykokou S, Ioannidis C. Image matching: a comprehensive overview of conventional and learning-based methods. *Encyclopedia* 2025; 5(1): 4. doi:10.3390/encyclopedia5010004.
- [2] Zhang S, Wang J, Li S, et al. Deep learning reforms image matching: a survey and outlook. *arXiv preprint* 2025; arXiv:2506.04619. doi:10.48550/arXiv.2506.04619.
- [3] Aharchi M, Ait Kbir M. A review on 3D reconstruction techniques from 2D images. In: *Proc 3rd Int Conf Smart City Appl (SCA 2020)*; 2020. pp. 510-522. doi:10.1007/978-3-030-37629-1_37.
- [4] Campos C, Elvira R, Rodríguez JG, Montiel JMM, Tardós JD. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Trans Robot* 2021; 37(6): 1874-1890. doi:10.1109/TRO.2021.3075644.
- [5] Chen H, Ye T, Li C, et al. Bright: a globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response. *Earth Syst Sci Data* 2025; 17(12): 6217-6253. doi:10.5194/essd-17-6217-2025.
- [6] Liu C, Zhang Y, Wang X, et al. A review of optical and SAR image deep feature fusion in semantic segmentation. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2024; 17: 12910-12930. doi:10.1109/JSTARS.2024.3424831.
- [7] Klaus B, Horn P. *Robot vision*. Cambridge, MA: MIT Press; 1986. ISBN: 978-0-262-08159-7.
- [8] Astruc G, Gonthier N, Mallet C, Landrieu L. AnySat: one Earth observation model for many resolutions, scales, and modalities. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR)*; 2025. pp. 19530-19540.
- [9] Jakubik J, Felsberg M, et al. Terramind: large-scale generative multimodality for Earth observation. In: *Proc IEEE/CVF Int Conf Comput Vis (ICCV)*; 2025. pp. 7383-7394.
- [10] Kang Q, Zhang J, Huang G, Liu F. Robust optical and SAR image matching via attention-guided structural encoding and confidence-aware filtering. *Remote Sens* 2025; 17(14): 2501. doi:10.3390/rs17142501.
- [11] Liao Y, Zhang X, Wang Z, et al. Feature matching and position matching between optical and SAR with local deep feature descriptor. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2022; 15: 448-462. doi:10.1109/JSTARS.2021.3134676.
- [12] Zhang Y, Qui C, Li S, et al. PLISA: an optical-SAR remote sensing image registration method based on pseudo-label learning and interactive spatial attention. *Remote Sens* 2025; 17(21): 3571. doi:10.3390/rs17213571.
- [13] Liao J, Yao Y, Yuan L, Hua G, Kang SB. Visual attribute transfer through deep image analogy. *ACM Trans Graph* 2017; 36(4): 120-1–120-15. doi:10.1145/3072959.3073683.
- [14] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)*; 2017. pp. 1125-1134.
- [15] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proc IEEE Int Conf Comput Vis (ICCV)*; 2017. pp. 2223-2232. doi:10.1109/ICCV.2017.244.
- [16] Quan D, Li X, Yang J, et al. Deep generative matching network for optical and SAR image registration. In: *Proc IEEE Int Geosci Remote Sens Symp (IGARSS)*; 2018. pp. 6215-6218. doi:10.1109/IGARSS.2018.8518653.
- [17] Du W, Zhou Y, Zhao J, Tian X. K-means clustering guided generative adversarial networks for SAR-optical image matching. *IEEE Access* 2020; 8: 217554-217572. doi:10.1109/ACCESS.2020.3042213.

- [18] Ma T, Ma J, Yu K, Zhang J, Fu W. Multispectral remote sensing image matching via image transfer by regularized conditional generative adversarial networks and local feature. *IEEE Geosci Remote Sens Lett* 2021; 18(2): 351-355. doi:10.1109/LGRS.2020.2972361.
- [19] Edstedt J, Sun Q, Bökman G, Wadenbäck M, Felsberg M. RoMa: robust dense feature matching. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR)*; 2024. pp. 19790-19800. doi:10.1109/CVPR52733.2024.01871.
- [20] Li L, Han L, Ye Y, Xiang Y, Zhang T. Deep learning in remote sensing image matching: a survey. *ISPRS J Photogramm Remote Sens* 2025; 225: 88-112. doi:10.1016/j.isprsjprs.2025.04.001.
- [21] Sommervold O, Gazzea M, Arghandeh R. A survey on SAR and optical satellite image registration. *Remote Sens* 2023; 15(3): 850. doi:10.3390/rs15030850.
- [22] Chen HM, Arora MK, Varshney PK. Mutual information-based image registration for remote sensing data. *Int J Remote Sens* 2003; 24(18): 3701-3706. doi:10.1080/0143116031000117047.
- [23] Wang Y, Yu Q, Yu W. An improved normalized cross correlation algorithm for SAR image registration. In: *Proc IEEE Int Geosci Remote Sens Symp (IGARSS)*; 2012. pp. 2086-2089. doi:10.1109/IGARSS.2012.6350961.
- [24] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)*; 2005; 1: 886-893. doi:10.1109/CVPR.2005.177.
- [25] Ye Y, Bruzzone L, Shan J, Bovolo F, Zhu Q. Fast and robust matching for multimodal remote sensing image registration. *IEEE Trans Geosci Remote Sens* 2019; 57(11): 9059-9070. doi:10.1109/TGRS.2019.2924684.
- [26] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981; 24(6): 381-395. doi:10.1145/358669.358692.
- [27] Lowe DG. Object recognition from local scale-invariant features. In: *Proc 7th IEEE Int Conf Comput Vis (ICCV)*; 1999; 2: 1150-1157. doi:10.1109/ICCV.1999.790410.
- [28] Xiang Y, Wang F, You H. OS-SIFT: a robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas. *IEEE Trans Geosci Remote Sens* 2018; 56(6): 3078-3090. doi:10.1109/TGRS.2018.2790483.
- [29] Ma W, Wen Y, Wu Y, et al. Remote sensing image registration with modified SIFT and enhanced feature matching. *IEEE Geosci Remote Sens Lett* 2017; 14(1): 3-7. doi:10.1109/LGRS.2016.2600858.
- [30] Li J, Hu Q, Ai M. RIFT: multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Trans Image Process* 2020; 29: 3296-3310. doi:10.1109/TIP.2019.2959244.
- [31] Li J, Shi P, Hu Q, Zhang Y. RIFT2: speeding-up RIFT with a new rotation-invariance technique. *arXiv preprint* 2023; arXiv:2303.00319. doi:10.48550/arXiv.2303.00319.
- [32] Zhao J, Yang D, Li Y, Xiao P, Yang J. Intelligent matching method for heterogeneous remote sensing images based on style transfer. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2022; 15: 6723-6731. doi:10.1109/JSTARS.2022.3197748.
- [33] Chen J, Zhang Y, Li S, et al. SAR and optical image registration based on deep learning with co-attention matching module. *Remote Sens* 2023; 15(15): 3879. doi:10.3390/rs15153879.
- [34] Borisov AN, Myasnikov VV, Sergeev VV. Method of automatic coregistration of digital remote sensing images from different sources. *Computer Optics* 2024; 48(6): 932-943. doi:10.18287/2412-6179-CO-1604.
- [35] Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag* 1901; 2(11): 559-572. doi:10.1080/14786440109462720.
- [36] Hu C, Zhu R, Sun X, Li X, Xiang D. Optical and SAR image registration based on pseudo-SAR image generation strategy. *Remote Sens* 2023; 15(14): 3528. doi:10.3390/rs15143528.
- [37] Dellinger F, Delon J, Gousseau Y, Michel J, Tupin F. SAR-SIFT: a SIFT-like algorithm for SAR images. *IEEE Trans Geosci Remote Sens* 2015; 53(1): 453-466. doi:10.1109/TGRS.2014.2323552.
- [38] Liaghat A, Helfroush MS, Norouzi J, Danyali H, Guerrero JM. OS-CycleGAN: modified CycleGAN-based descriptors for optical and SAR image matching. *IEEE Trans Geosci Remote Sens* 2025; 63: 5214918. doi:10.1109/TGRS.2025.3575193.
- [39] Wu B, Wang H, Zhang C, Chen J. Optical-to-SAR translation based on CDA-GAN for high-quality training sample generation for ship detection in SAR amplitude images. *Remote Sens* 2024; 16(16): 3001. doi:10.3390/rs16163001.
- [40] Sghaier MO, Hadzagic M, Yu JY, Shton S. Change detection in heterogeneous satellite remote-sensing imagery using deep features comparison. In: *Proc IEEE Int Geosci Remote Sens Symp (IGARSS)*; 2024. pp. 8599-8602. doi:10.1109/IGARSS53475.2024.10641813.
- [41] Du Z, Zhang Y, Li M, et al. Concatenated deep-learning framework for multitask change detection of optical and SAR images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2024; 17: 719-731. doi:10.1109/JSTARS.2023.3333959.
- [42] Schmitt M, Hughes LH, Zhu XX. The SEN1-2 dataset for deep learning in SAR-optical data fusion. *ISPRS Ann Photogramm Remote Sens Spatial Inf Sci* 2018; IV(1): 141-146. doi:10.5194/isprs-annals-IV-1-141-2018.
- [43] Schmitt M, Hughes LH, Qui C, Zhu XX. SEN12MS—a curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion. *ISPRS Ann Photogramm Remote Sens Spatial Inf Sci* 2019; IV(2/W7): 153-160. doi:10.5194/isprs-annals-IV-2-W7-153-2019.
- [44] Xiang Y, Tao R, Wang F, You H, Han B. Automatic registration of optical and SAR images via improved phase congruency model. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2020; 13: 5847-5861. doi:10.1109/JSTARS.2020.3026162.
- [45] Wenger R, Puissant A, Weber J, Idoumghar L, Forestier G. MultiSenGe: a multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications. *ISPRS Ann Photogramm Remote Sens Spatial Inf Sci* 2022; V-3-2022: 635-640. doi:10.5194/isprs-annals-V-3-2022-635-2022.
- [46] Borisov A. Methods of spectral matching for remote sensing data. *Opt Mem Neural Netw* 2025; 34(Suppl 2): S221-S229. doi:10.3103/S1060992X25602398.
- [47] Borisov AN, Myasnikov VV, Sergeev VV. Methods of spectral matching of heterogeneous remote sensing images. *Optoelectron Instrum Proc* 2025; 61(4): 422-429. doi:10.3103/S8756699025700475.
- [48] Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging* 2017; 3(1): 47-57. doi:10.1109/TCI.2016.2644865.

- [49] Zhou W, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004; 13(4): 600-612. doi:10.1109/TIP.2003.819861.
- [50] Kramer MA. Autoassociative neural networks. Comput Chem Eng 1992; 16(4): 313-328. doi:10.1016/0098-1354(92)80051-A.
- [51] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proc 18th Int Conf Med Image Comput Comput Assist Interv (MICCAI); 2015. pp. 234-241. doi:10.1007/978-3-319-24574-4_28.
- [52] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv preprint 2019; arXiv:1905.11946. doi:10.48550/arXiv.1905.11946.
- [53] Edstedt J, Bökman G, Zhao Z. DeDeDo v2: analyzing and improving the DeDeDo keypoint detector. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops (CVPRW); 2024. pp. 4245-4253. doi:10.1109/CVPRW63382.2024.00428.
- [54] Iakubovskii P. Segmentation Models PyTorch. 2020. Available at: https://github.com/qubvel-org/segmentation_models.pytorch.
- [55] Dozat T. Incorporating Nesterov momentum into Adam. 2016. Available at: <https://openreview.net/forum?id=OM0jvwB8Jlp57ZJtNEZ>.
- [56] Smith LN, Topin N. Super-convergence: very fast training of neural networks using large learning rates. In: Proc SPIE 2019; 11006: 369-386. doi:10.1117/12.2520589.
- [57] Zhu JY. CycleGAN and pix2pix in PyTorch. Available at: <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>.

Сведения об авторах

Борисов Алексей Николаевич, 1996 года рождения, в 2019 г. окончил с отличием Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет). В настоящее время работает старшим преподавателем кафедры геоинформатики и информационной безопасности Самарского университета. Круг научных интересов включает высокопроизводительные вычисления, вычисления общего назначения на графических процессорах, обработку изображений. E-mail: borisov.an@ssau.ru

Мясников Владислав Валерьевич, в 2008 защитил диссертацию на соискание степени доктора физико-математических наук. Работает профессором кафедры геоинформатики и информационной безопасности Самарского национального исследовательского университета имени академика С.П. Королева. Круг научных интересов: компьютерное зрение, машинное обучение, искусственный интеллект и геоинформатика. Имеет более 200 публикаций, в том числе более 100 статей и три монографии (в соавторстве). E-mail: vmyas@geosamara.ru

Сергеев Владислав Викторович, 1951 года рождения. В 1974 году окончил Куйбышевский авиационный институт (ныне Самарский национальный исследовательский университет имени академика С. П. Королёва). В 1978 г. защитил кандидатскую, а в 1993 году – докторскую диссертацию. В настоящее время является профессором кафедры геоинформатики и информационной безопасности Самарского университета. Область научных интересов: цифровая обработка сигналов, анализ изображений, распознавание образов, геоинформатика. E-mail: vserg@geosamara.ru

Поступила в редакцию 15 января 2026 г. Окончательный вариант – 25 февраля 2026 г.

Common domain construction method for radar and optical Earth remote sensing images matching

A.N. Borisov¹, V.V. Myasnikov¹, V.V. Sergeev¹

¹ Samara National Research University, Moskovskoye Shosse 34, Samara, 443086, Russia

Abstract

This paper presents a new method for constructing a common domain for digital Earth remote sensing images acquired by two types of sensors: a multispectral optical sensor and a synthetic aperture radar. After translation to a common domain, initially heterogeneous images become sufficiently similar to be coregistered by image matching methods, developed for standard images. The problem of constructing a common domain is considered as a task of simultaneous training of two U-Net neural networks. A proposed method is compared with an existing method based on the pix2pix generative neural network, which performs domain translation from the radar to the optical domain. Results of image matching using the well-known SIFT algorithm, the state-of-the-art RoMa and DeDoDe algorithms, and the RIFT2 heterogeneous image matching algorithm are presented. Experimental studies demonstrate the clear advantage of a proposed method: a proposed method in combination with the trained RoMa network achieves the matching success rate of 99.9% with the average corner error of 0.84 pixels; pix2pix in combination with DeDoDe achieves the matching success rate of 98% with the average corner error of 4.3 pixels.

Keywords: digital Earth remote sensing images, multispectral images, radar images, common domain, image coregistration.

Citation: Borisov AN, Myasnikov VV, Sergeev VV. Common domain construction method for radar and optical Earth remote sensing images matching. *Computer Optics* 2026; 50(3): 1980. DOI: 10.18287/COJ1980.

Acknowledgements: This work was supported by the Ministry of Science and Higher Education, Russia (Agreement No. 075-15-2024-558).

Author's information

Alexey Nikolayevich Borisov, (b. 1996), graduated with honors from Samara National Research University (Samara University) in 2019. Currently works as a senior lecturer at the Geoinformatics and Information Security department at Samara University. Research interests: high-performance computing, GPGPU, image processing. E-mail: borisov.an@ssau.ru

Vladislav Valerievich Myasnikov, received his DrSc degree in Physics & Maths (2008). Currently he works as a professor at the Geoinformatics and Information Security department in Samara National Research University. The range of scientific interests: computer vision, pattern recognition and artificial intelligence, machine learning and geoinformatics. He has about 200 publications, including more than 100 articles and three monographs. E-mail: vmyas@geosamara.ru

Vladislav Viktorovich Sergeev, (b. 1951) graduated from Samara National Research University (Samara University) in 1974. In 1978 he received the Candidates of Technical Sciences degree and in 1993 – the Doctor of Technical Sciences degree. Currently he works as a professor at the Geoinformatics and Information Security department in Samara National Research University. Research interests are digital signals and image processing, GIS and pattern recognition. E-mail: vserg@geosamara.ru

Received January 15, 2026. The final version – February 25, 2026.
